DOCUMENT RESUME

ED 218 282 TM 820 194

AUTHOR Guyette, Susan; Churchman, David
TITLE Two Models for Ethnographic Evaluation
PUB DATE Dec 81
NOTE 18p.; Paper presented at the Annual Meeting of the American Anthropological Association (Los Angeles, CA, December, 1981).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Decision Making; *Educational Anthropology; *Ethnography; *Evaluation Methods; Evaluators; *Models; *Program Evaluation; Social Science Research
IDENTIFIERS Evaluation Research; *Qualitative Analysis

ABSTRACT

An ethnographic evaluation adapts data collection and analytical methods of the ethnographer to the purpose of decision-making. Ethnography and evaluation differ in determination of problems addressed by the researcher. Ethnographers are interested in slow cultural change, while evaluators address policy questions requiring immediate answers. The micro-ethnographic evaluation limits the scope of the study to the school, viewed as a subculture. The resultant data is limited to the perception of a small number of participants. The macro-ethnographic evaluation considers school interaction as interrelated with other cultural subsystems. The process of enculturation occurs as part of the curriculum. The advantages and disadvantages of micro-and macro-ethnographic evaluations are outlined. The ethnographic approach to evaluation reduces the cultural bias of evaluators, thus making it appropriate in explaining programs for minorities. Another advantage of this approach is eliminating the necessity for a control group. It does not require statistical methods, thus making evaluation reports more easily understandable. The cultural suitability of the educational program is assessed through a macro-ethnographic evaluation. Ethnographic evaluations are appropriate when a comprehensive understanding of the program and its relationship to society is the major factor in judging program quality. (DWH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Churchman, D.
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED218282

TWO MODELS FOR ETHNOGRAPHIC EVALUATION

Susan Guyette (University of California, Los Angeles)
David Churchman (California State University, Dominguez Hills)

I. INTRODUCTION

In recent years, educational anthropologists have turned attention toward the techniques of program evaluation. The purpose of evaluation is to measure progress toward the accomplishment of goals. As an applied discipline, evaluation ultimately assists educators and administrators to make decisions regarding program improvements or to become aware of program results, thus enabling them to make informed judgements about program results. During the past decade dissappointment with evaluations often focused upon the tendency toward cultural bias exhibited by the more frequently dominant-culture evaluators (Sjoberg 1975:43). That is, the formulation of evaluation questions, the interpretation of goals, and methods of measuring progress are particularly susceptible to cultural skew. With this particular problem in mind, educators and administrators have with increasing frequency, called on anthropologists to participate in this form of contract research.

The majority of ethnographic evaluations conducted thusfar by educational anthropologists are classroom ethnographies, which focus on the process of classsroom interaction. Although this emphasis answers important questions concerning what actually happens in the classroom setting, it is limited for decision-making applications. Use of the wide range of evaluative techniques can broaden the kind of results suitable for

TM 820 154

Presented at the annual meeting of the American Anthropological Association, Los Angeles, CA, December 1981

decision-making situations. The purpose of this paper is to discuss six issues in evaluation design, while integrating evaluation and ethnographic methodology appropriate for the applied setting. Then, two models are presented for ethnographic evaluation. As an introduction, we will discuss a brief history of evaluation research, outlining the successes and failures of past techniques. It is from this point that new research paradigms can be developed for the culturally sensitive evaluation.

II. HISTORY AND LIMITATIONS OF EVALUATION

Program evaluation became important, in terms of resources utilized, responsibilities undertaken and results expected, as a consequence of the social program experiments initiated during the "Great Society" era of the 60's. Essentially, program evaluation was the method by which the Congress expected to maintain accountability, assess results and identify successful programs worthy of continuation and imitation.

The initial attempts to carry out these responsibilities relied on traditional research design descended from the methods originated by Ronald Fisher in the late nineteenth and early twentieth centures. These methods relied on randomization, control groups and statistical analysis of data for the purpose of inferring causal relations between independent and dependent variables. They had been progressively refined through subsequent experimental work, primarily in university psychology laboratories, and had recently become capable of handling large amounts of data with the advent of the computer.

The educational programs initiated during the Great Society, more than most of the other experiments of the period, had the

additional advantage of an established tradition for measuring effects. This tradition was based on the work of Edward Thorndike, Lewis Terman, James Cattell and others. The emphasis was on the use of paper and pencil tests administered to individual students and summarized for individuals, classes and higher groups through descriptive statistics.

Evaluation based on these traditions required random assignment of students to an experimental program or usually to an existing educational program as a control; testing of students to generate data at predetermined points in time; and statistical tests to determine the relative success of the two programs. Virtually all the evaluations done for the following decade represented variations on this basic scheme.

Results were disappointing. Random assignment frequently proved impossible in school settings, and in almost all of the social programs of the period. Tests used to collect data were predisposed to favor either the experimental or the control group, making interpretation of results unexpectedly difficult. Insufficient attention was given to verifying the extent to which the experimental program actually was carried out, leaving ambiguous exactly what had to be done to achieve the same results again. Halo effects, Hawthorne effects[1] and contamination of the experimental program by the control and vice-versa further complicated interpretation of results. Knowledge that an experimental program was more successful than the existing program at some particular school was inadequate in making a decision about the experimental program unless the regular program at that school was identical to the control group program; even if the experiment went well, the evaluation did not provide very useful information. But, it did overwhelm people untrained in the traditions of Fisher and Thorndike with

the appearance of accuracy and objectivity, a consequence of the tables and numbers to several decimal points, which dominate the reports. Tests also often were found to be culturally biased when used to measure progress of minority students who were the major recipients of Great Society assistance.

More important than these technical difficulties, the administrators for whom these evaluations were done often were dissatisfied with the results. The typical evaluation described above, carried out at great expense, usually may be boiled down to the single conclusion that there is, or is not, a statistically significant difference between two educational programs. Typically, the educational significance of this statistical difference is not addressed because comparable "objective" standards do not exist and the conclusion would have to be reached "subjectively", anathema to the tradition; yet this is the most important question for the administrator. The approach described above leaves totally unaddresssed other equally vital questions, such as the costs of the experimental program, or the way different groups in the community would react to its adoption, or the relationship of the experimental program to the school curriculum as a whole.

Difficulties such as these with traditional methods for conducting evaluations are among the reasons many have argued that broader, richer methods must be substituted. Among the many that have been suggested is ethnography, which uses a wide range of both quantitative and qualitative methods.

III. DIFFERENCES BETWEEN ETHNOGRAPHY AND EVALUATION

There are some important differences between ethnography and evaluation that should be recognized before the two disciplines

can be integrated. The most important of these differences is in determining the problem that is addressed by the researcher. Ethnographers traditionally are interested in culture, which changes slowly; evaluators address policy questions that typically require answers without delay. Ethnographers, under little or no time pressure, typically allow the issues to be studied to emerge over relatively long periods of time. Evaluators not only work to specific short-term deadlines, but must address questions raised by someone else. Thus, ethnographers expect to learn the nuances and meanings of a cultural system after years spent gaining rapport with people in the culture being studied. Evaluators get in, get the best information they can and get out quickly. Ethnographers are process-oriented, while evaluators concentrate on outcomes.

One consequence of these differences is that ethnographers often feel that their science is being exploited by evaluators who are using their techniques in new ways for new purposes. But, man has a long history of modifying existing tools to achieve new purposes. Adapting ethnographic methods to evaluation studies requires selection of the relevant elements of traditional ethnographic methods and ignoring others not relevant to evaluation tasks. The plethora of names such as "survey ethnography" and "contract ethnography" may represent recognition that adaptation rather than adoption is taking place.

In this view, the fundamental question is whether the traditional methods of ethnographic data collection and analysis can be adapted to conducting an evaluation that better meets the needs of a program administrator than does the more common experimental method. Traditionally, ethnographers make inferences from what people say, what people do and what people

use. They commonly rely on basic tools such as the interview, the questionnaire and observation. Evaluators employ these basic tools in addition to ratings or rankings (by staff, participants, or managers), journals kept by staff, case histories, organizational record keeping, analysis of documents, clinical examinations, information tests, standardized attitude inventories, and ability tests. Although the ethnographic evaluation efforts of the past have utilized primarily the former, and the experimental design draws primarily from the latter techniques, combinations of qualitative and quantitative techniques may provide the basis for future ethnographic evaluations which both meet decision-making needs and are culturally sensitive.

IV. ISSUES IN EVALUATION DESIGN

Design of an evaluation involves a large number of often contradictory considerations which interact with one another. The inevitable consequence is that tradeoffs must be made among design factors and compromises results in less than perfect studies. But, it is vital that the evaluator understand what these tradeoffs and compromises are so that the best possible evaluation is provided in a specific circumstance. This section of the paper discusses six such design issues.

A first design decision concerns the audience for whom the evaluation is intended. The audience for an evaluation consists of decisonmakers who face a specific deadline. These decisionmakers may be divided into those "internal " to the organization running the program being evaluated (e.g. program staff or administrators), and those "external" to the organization (for example, a funding agency or legislative

body). The former usually want help in improving their program and some evidence that they are doing a good job. The latter usually want to know such things as whether the program should continue to receive funds and if so at what level. The term "formative evaluation" often is used to distinguish the type of evaluation done for staff and administrators, while the term "summative evaluation" is used for that aimed at funding agencies or legislators. We will discuss these terms below more fully. Internal evaluation has the advantage of a much greater understanding of organizational dynamics, history, personalities, capabilities and the like, and therefore often results in a more sensitive evaluation. On the other side of the coin, unless the organization has evaluation specialists available, the evaluation may not be completed properly or on time and the results may be less credible to external audiences. External evaluation has almost precisely the opposite advantages and disadvantages. As a general rule, formative evaluations can be conducted internally, but summative evaluations should be conducted by an outsider.

A **second** issue concerns the application of the evaluation, for purposes of **decision** or **judgement**. Ethnographers are expected to avoid making a judgement about the society studied; even less acceptable might be the idea of offering suggestions as to how to make improvements. Having accepted this philosophy, they may have some difficulty accepting the applied character of evaluation research, which is aimed, ultimately, at judgement and decisionmaking. Within evaluation circles, the debate is whether the evaluator or the employer should draw conclusions, and whether these conclusions should be limited to practical decisions or to more comprehensive judgements. It is a perfectly acceptable position to insist that you, as an

evaluator, will not include recommendations or judgements in your evaluation, but it would be inappropriate to accept a contract to provide an employer with information needed to make judgements or decisions and not to provide that information. Further, it is important that you understand the purposes for which the evaluation are to be undertaken before accepting the contract or designing the evaluation. As suggested, these fall into the general categories of (1) provision of information required to make specific decisions and (2) provision of information required to make judgements about the program.

A third choice facing the evaluator during the formulation of the evaluation design concerns the **evaluation question.** This step in problem definition is perhaps one most likely to be influenced by cultural perception. An evaluation that will meet the information needs of an educational administrator may provide answers to evaluation questions[2] such as :

* What work is done in the program being evaluated--that is, how do students and teachers actually spend their time?
* How is "individualized instruction" experienced by students? (versus how well does individualized instruction work?)
* How do the various types of work done in the program relate to one another?
* What are the major stages through which the program moves, and what work is done within each stage?
* How is achievement measured or identified and what status or rewards does it provide?
* How is the program administered or controlled?

* How does the program respond to the cultural differences or similarities of the culture-at-large?

* How do the activities in the program relate to the cultural activities in the community?

As the questions asked during any evaluation will depend on the decisions that have to be made, this list is illustrative, suggestive and incomplete. The task of the evaluator is to help define the necessary questions, then to collect the necessary information and to analyze it. The actual decisions are necessarily the task of the administrator. This is consistent with definitions that view the evaluator as a collector, analyzer and interpreter of information, and recognize that the authority for decisions must remain with the administrator who is responsible for the program (e.g. Alkin, 1969).

A **fourth** design issue concerns type of evaluation appropriate to the evaluation question. One of the most fundamental theoretical distinctions made by evaluators is that between **formative** and **summative** evaluation (Scriven, 1967). The purpose of the former is improvement; that of the latter is certification or validation. Formative evaluation implies identification of elements of an ongoing program that are not working as well as they might, so that improvements can be made. Formative evaluation is an ongoing process, throughout the development of a project. Data are collected while the program is in process, to determine the continual effectiveness of the program and make changes in the program activities if necessary. Using this approach can save resources and improve results, by "forming" the program as effectiveness is determined. Formative evaluation should, logically, be carried on until the cost of conducting the evaluation exceeds the value of improvements that

can be made in the program. Summative evaluation occurs after the program is final or completed, although pretesting of the data collection instruments and the data collection usually occurs while the program is in progress. The program then should be carried out in a homeostatic or unchanging state. The results of a summative evaluation are often used in making decisions as to keeping a particular program structure for future use. The summative evaluation is particularly useful in comparing different types of programs and deciding which type is the most effective. A summative evaluation often enables program administrators and funders to decide whether the program should be adopted permanently or in some wider context.

The theoretical distinction is fairly clear, but in practice many evaluations claim to be one (for example, summative) but include information relevant to the other (for example, recommendations on how to improve a progam). One reason for this state of affairs is the extent to which evaluators rely on traditional research design. Logically, different questions require different methods, but this has been rare in practice. The distinction between formative and summative evaluation should be kept in mind in selecting the methodology for a particular evaluation problem.

A fifth issue addresses the methods of goal-based versus goal-free evaluations. The overwhelming majority of evaluations are conducted by measuring the extent to which the intended objectives of the project are being achieved, but some writers have argued that the goal-based method gives the evaluator "tunnel vision" that leads to overlooking important side effects. These effects could be desirable or not. For example, a project to teach children to read may awaken their interests in whatever the subject matter of the reading material is ---

endangered species or computers or art history or whatever. Usually this would be considered a desirable unintended effect. If it were missed by an evaluator who only measured reading skills, then the value of the program would be underestimated.

The goal-free method differs from the goal-based in that the objectives are not known by the evaluator at the start of the evaluation. They are discovered or elicted from participants during the evaluative process. Goal-free evaluation has potential for use with ethnographic methods such as participant-observation and domain analysis. Also, the goal-free approach may be more appropriate in minority communities where the implemented goals may differ from those defined by dominant culture educators. Whenever unintended effects might significantly influence the decisions to be made about the project being evaluated and there is reason to suspect that the effects cannot be predicted, goal-free evaluation should be considered, either as an alternative or as a supplement to the more traditional goal-based evaluation. However, the choice should not be made lightly. By their very nature unintended effects are difficult to detect, particularly when it is realized that they may take place beyond the confines of the project. Further, there are no generally accepted methods for conducting goal-free evaluation, and participants may perceive the process as very threatening because they have no idea of what the evaluator is looking for.

A sixth issue involves decisions regarding data. Throughout the heyday of evaluations conducted of Great Society programs, there was a distinct preference for quantitative data. It was seen as more reliable and thus more valid. It was more "objective". More sophisticated and sensitive statistical techniques were available for analysis of quantitative data.

But, results did not live up to expectations. Gradually, criticism grew. Reliability and validity were gained by measuring variables that were easy rather than important to measure. Objectivity was compromised due to factors such as cultural bias. The inability of refined techniques to find differences between radically different programs suggested that statistical and educational significance were quite different from one another. Led by Robert Stake, the statistical evaluators began to question the appropriateness of their methodological assumptions, and then to advocate so-called qualitative methods. Advantages of the qualitative approach are that it can be discovery-oriented, exploratory, descriptive, inductive, process-oriented, and holistic. More recently evaluators and ethnographers have adapted qualitative techniques for ethnographic evaluations. The consequence has been to vastly extend methods considered acceptable for conducting evaluations. Now, instead of blindly choosing quantitative data gained through a traditional research design, the evaluator must make a rational choice from among many possibilities and defend the choice against all others as the most appropriate for a particular circumstance.

V. TWO MODELS FOR ETHNOGRAPHIC EVALUATION

To the present, educational anthropologists in applied settings usually adapt evaluation techniques while conducting classroom ethnography. The potential use of evaluative techniques extends beyond this scope, and to illustrate an alternative model for ethnographic evaluation the distinction between micro-and macro- ethnographic evaluation is developed.

The micro- ethnographic evaluation limits the scope of the study to the school. Supporters of this approach (Wilson 1977) view the school as a subculture, utilizing ethnographic methods to analyze and interpret activities according to the participant's reality. The resulting product of this evaluation is generally an ethnography describing the program activities, who interacted with whom, and the participants' interpretations. The product may address the question, "did the program accomplish its goals?" by discovering the goals within the school considered as a closed system. One common accomplishment of the school ethnography is the possible discovery of discrepancies between the ideal or stated behavior and the real, or actual behavior as reflected in activities. Through ethnographic interview within the school, it is possible to ask participants if activities in the school are connected to elements of the culture as a whole. This data is limited, however, to the perception of a small number of participants.

In contrast, the macro-approach to ethnographic evaluation considers school interaction as interrelated with other cultural subsystems (Wolcott 1975; Everhart 1975). In other words, the school is not viewed as a closed system, but rather as a part of the culture as a whole. This approach takes into account the process of enculturation, whereby the transmission of that cultural knowledge shared by the community occurs as part of the curriculum. There are three major steps to conducting the macro analysis. First, a study is conducted to determine how each of the cultural subsystems work in the community. This involves observing, participating, and interviewing to identify activities. The cultural meaning of activities is identified , and domains of interaction are identified for the culture. Second, a domain

analysis is conducted with data obtained from activities within the school. And thirdly, the final analysis compares the domains and world view of the culture-at-large with the transmission of knowledge in the school. Similarities and differences are then identified through this comparison. When disjunctions exist, the more common consequence is community opinion that the school does not reflect the social processes of the culture.

Advantages of the micro-approach over the macro-approach include lowered costs, reduced time, and narrowly defined boundaries. Disadvantages of the micro-approach are that the study is more likely to be biased by the goals set by the educators and the scope of the study is limited. The major, and extremely important, advantage of the macro-approach is the discovery of discrepancies between the educational goals of the community and the actual activities within the school. For this reason, the macro-approach is more suitable to non-Western cultures in particular. Where time and funds may pose constraints upon the macro-approach, existing ethnographies can supplement ethnographic fieldwork in the culture-at-large.

There is an important advantage of the macro-approach over the micro-approach that should be considered. With the micro-approach, if the school is studied as a subculture, then connections to the culture-at-large are minimal. In the macro-approach, when the culture is examined first, followed by an analysis of the activities in the school, two types of information are discovered. It is possible to see if there are no connections between activities in the culture and the activities in the schools, as well as if the activities in the

schools do not connect to the culture. This information gives a basis for determining if items need to be added to the curriculum to reflect needed cultural learning, or if there are items in the curriculum that need to be deleted. The purposes of the evaluation, advantages, and disadvantages of each approach should be weighed in deciding upon the more appropriate ethnographic evaluation.

VI. SUMMARY

An ethnographic evaluation adapts the data collection and analytical methods of the ethnographer to the purpose of decision-making. The two most important of these adaptations are advance specification of the questions to be addressed and specific, usually eminent, deadlines that preclude the development of rapport usually desired by ethnographers. Such a summative evaluation remains faithful to the traditions of ethnography in attempting to understand educational programs from the viewpoint of the participants. Interpretations probably will emphasize description of program progress and structure rather than how much course content has been absorbed by students or what their attitudes are toward the subject. One advantage of taking the ethnographic approach is in reducing the cultural baggage, or the values that the evaluator carries with him into the cultural setting. These values can introduce bias into the formulation of the evaluation questions or into the final recommendations. The approach presented is therefore particularly appropriate in explaining programs for minorities within any culture.

The ethnographic approach has the second advantage of not requiring a control group. Judgements can be reached without

having to set up or explain to parents why children are receiving different programs, or in situations where the total number of students does not permit such division. If the purely qualitative methods we have suggested are the only ones used, the judgement must be made without the achievement data characteristic of most evaluations. This may not be the weakness it appears. First, previous work suggests that one can estimate program effects just as well from time on task as from tests. Second, testing poses special problems of cultural bias particularly for minority students. Third, the statistically significant results reported from control-group evaluations often seem too small to be educationally significant, so have little impact on the overall judgement made about a program.

A third advantage of the qualitative ethnographic approach we have suggested is that it does not require the often esoteric statistical methods. The result is an evaluation report that the administrator who seldom understands these methods does not have to take on faith. Thus, the reports themselves are more useful to the people for whom they are intended.

A fourth advantage of the ethnographic approach is that the activities and views of the culture-at-large are linked to the activities within the school. This enables the discoveries of linkages and disjunctions in the educational process, providing direction for revision of curriculum content. The cultural appropriateness of the educational program is assessed through use of the macro-ethnographic evaluation.

It appears then that ethnographic methods may be useful for the task of conducting evaluations. They are particularly suitable when a comprehensive understanding of the program and its relationship to society is the most important factor in judging program quality, as opposed to such other bases as cost

or cognitive achievement. But, the decision-making purpose of evaluative research requires adaptation rather than adoption of traditional ethnographic methods.

NOTES

1. Hawthorne effects concern bias introduced by the subject through knowledge of inclusion in the experiment; whereas, Halo effects are introduced when the researcher or teacher is biased due to prior impressions of the student.

2. "Evaluation questions" define the purpose of the study; "interview questions" are asked of participants selected from the program being evaluated.

REFERENCES

Alkin, M.C. "Evaluation Theory Development", Evaluation Comment, 1969, 2(1), pp. 2-7.

Everhart, R. B. "Problems of Doing Fieldwork in Educational Evaluation", Human Organization, 1975, 34(2), pp. 205-215.

Scriven, M. "The Methodology of Evaluation", in Perspectives of Curriculum Evaluation, Eds. R.W. Tylor, R.M. Gange, and M. Scriven, (Chicago: Rand McNally & Co., 1967).

Sjoberg, Gideon "Politics, Ethics and Evaluation Research", in Handbook of Evaluation Research, Vo. 2, 1975, pp. 29-54.

Wilson, S. "The Use of Ethnographic Methods in Educational Evaluation", Human Organization, 1977, 36(2), pp. 200-203.

Wolcott, H. "Criteria for an Ethnographic Approach to Research in Schools", Human Organization, 1975, 34(2), pp. 111-127.